

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:46 : Search time 170.72 Seconds
(without alignments)
23.864 Million cell updates/sec

Searched: 195891 seqs, 6790655 residues

total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database : PIR_66;*

1: pir1;*

2: pir2;*

3: pir3;*

4: pir4;*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length	DB ID
1	98.5	28.7	B24810
2	93	27.1	FWSYBA
3	93	27.1	S20007
4	75	21.9	T29475
5	70	20.4	T31863
6	69	20.1	T18355
7	68.5	20.5	T20681
8	67.5	19.7	T31889
9	67.5	19.7	T31998
10	66.5	19.4	S08059
11	66.5	19.4	D49423
12	66	19.2	T29699
13	66	19.2	T31887
14	66	19.2	S06398
15	66	19.2	A36502
16	64	18.7	1300
17	63.5	18.5	FWCNAB
18	63.5	18.5	158169
19	63	18.4	T48477
20	63	18.4	T29880
21	63	18.4	S32991
22	62.5	18.2	G69186
23	61.5	17.9	PN0667
24	61	17.8	S22145
25	61	17.8	T03720
26	61	17.8	A39827
27	61	17.8	B39827
28	61	17.8	T20562
29	60.5	17.6	T27641

RESULT 1

B24810

C:Species: Glycine max (soybean)

C:Date: 07-Mar-1988 #sequence_revision 30-Jun-1991 #text_change 20-Aug-1999

C:Accession: B24810; S16337

R:Doyle, J.J.; Schuler, M.A.; Godette, W.D.; Zenger, V.; Beachy, R.N.; Slightom, J.L.; Biol. Chem. 261, 9228-9238, 1986

A:Title: The glycosylated seed storage proteins of Glycine max and Phaseolus vulgaris

A:Reference number: A24810; MUID:86250867

A:Accession: B24810

A:Molecule type: DNA

A:Residues: 1-639 <DOV>

A:Cross-references: GB:MI3759; NID:gi169928; PID:AAB01374_1; PID:gi169929

A:Note: the authors translatin the codon GGT for residue 352 as Glu

R:Schuler, M.A.; Schmitt, E.S.; Beachy, R.N.

Nucleic Acids Res. 10, 8223-8244, 1982

A:Title: Closely related families of genes code for the alpha and alpha' subunits of A:Reference number: S16337; MUID:83143288

A:Accession: S16337

A:Molecule type: DNA

A:Residues: 361-639 <SCH>

C:Genetics:

A:Translations: 278/1; 355/3; 382/3; 481/3; 575/1

C:Superfamily: glycinin

C:Keywords: seed; storage protein

Query Match 28.7%; Score 98.5; DB 2; Length 639;
Best Local Similarity 30.6%; Pred No. 0.0029;
Matches 22; Conservative 10; Mismatches 25; Indels 15; Gaps 3;

QY 2 ENPCAQRCLOSCQEQPDLKQAKCESRQT-KLEYDPRCVYDPRG-----ATNQR 48

Db 31 QPSHKNKCLRSCCNSEKDSDYRNQACHARCNLLKVEEEBC---ERGQIPRPRQPQHPEREROQ 88

QY 49 HPGGERTRGRQP 60

Db 89 HGERKEDEBEGEQP 100

RESULT 2

FWSYBA

C:Species: beta-conglycinin alpha chain precursor - soybean

C:Accession: S14481; S74124; S06714

C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999

C:Accession: S14481; S74124; S06714

R:Sebastiani, F.L.; Farrell, L.B.; Schuler, M.A.; Beachy, R.N.

Plant Mol. Biol. 15, 197-201, 1990

A:Title: Complete sequence of a cDNA of alpha subunit of soybean beta-conglycinin.

A:Reference number: S14681; MUID:9135860

A:Accession: S14481

A; Molecule type: mRNA
A; Residues: 1-605 <SEB>
A; Cross references: EMBL:X17698; NID:918535; PIDN:CAA35691.1; PRID:918536
R; Shutov, A. D.; Rakhovskaya, T. A.; Bastrygina, A. S.; Bulmaga, V. P.; Horstmann, C.; Muent
Eur; J. Biocn., 241, 221-228, 1996
A; Title: Limited proteolysis of beta-conglycinin and glycinin, the 7S and 11S storage g
A; Reference number: S74123; MUID:97054613
A; Accession: S74124
A; Molecule type: protein
A; Residues: 189-196, 'H', 198, 'N', 200, 'X', 202-203; 397-408, 'X', 410, 'X', 412-417, 'X', 419-420,
A; Experimental source: seed
C; Superfamily: glycinin
C; Keywords: glycoprotein; seed; storage protein
F; 1-22/Domain: signal sequence #status predicted <PRO>
F; 23-62/Domain: propeptide #status predicted <PRO>
F; 63-605/Product: bata-conglycinin alpha chain #status predicted <PRO>
F; 261,517/Binding site: carbohydrate (Asn) (covalent) #status predicted <PRO>
Query Match 27.1%; Score 93; DB 1; Length 605;
Best Local Similarity 31.4%; Pred. No. 0.011; 16; Indels 24; Gaps 3;
Matches 22; Conservative 8; Mismatches 16; Indels 24; Gaps 3;

Qy 2 ENPACQRQLQSCQEPDDLKQAKCESRTKLEYD-----PRCVYDGTGATNRP 50
Db 31 ENPKHNKLCQSCNSERDSYRNQACHARCNLLKVEKECEEGEIPRPR-----PRPQHP 83
Qy 51 -----PGER 54
Db 84 ERIFQQPGEK 93

RESULT 3
S20007
beta-conglycinin alpha chain precursor - soybean
C; Species: Glycine max (soybean)
C; Accession: S20007
C; Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Oct-1997
R; Believre, J.-M.; Dickinson, C.D.; Dickinson, L.A.; Nielsen, N.C.
Plant Mol. Biol. 18, 259-274, 1992
A; Title: Synthesis and assembly of soybean beta-conglycinin in vitro.
A; Reference number: S20007; MUID:92119248
A; Accession: S20007
A; Status: preliminary
A; Molecule type: mRNA
A; Residues: 1-605 <LEL>
C; Superfamily: glycinin
Query Match 27.1%; Score 93; DB 2; Length 605;
Best Local Similarity 31.4%; Pred. No. 0.011; 16; Indels 24; Gaps 3;
Matches 22; Conservative 8; Mismatches 16; Indels 24; Gaps 3;

Qy 2 ENPACQRQLQSCQEPDDLKQAKCESRTKLEYD-----PRCVYDGTGATNRP 50
Db 31 ENPKHNKLCQSCNSERDSYRNQACHARCNLLKVEKECEEGEIPRPR-----PRPQHP 83
Qy 51 -----PGER 54
Db 84 ERIFQQPGEK 93

RESULT 4
T29475
hypothetical protein T01D1.6 - Caenorhabditis elegans
C; Species: Caenorhabditis elegans
C; Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
R; Bradshaw, H.; Wohlgemann, P.
Submitted to the EMBL Data Library, November 1996
A; Description: The sequence of C. elegans cosmid T01D1.
A; Reference number: 220523
A; Accession: T29475

RESULT 5
T31863
hypothetical protein c54f6.6 - Caenorhabditis elegans
C; Species: Caenorhabditis elegans
C; Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C; Accession: T31863
R; Geisel, C.; Bradshaw, H.
submitted to the EMBL Data Library, July 1997
A; Description: The sequence of C. elegans cosmid C54F6.
A; Reference number: Z21094
A; Accession: T31863
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: DNA
A; Residues: 1-221 <GET>
A; Cross references: EMBL:AF016447; PIDN:AAB65937.1; GSPDB:GN00023; CESP:C54F6.6
A; Genetics:
A; Gene: CESP:C54F6.6
A; Map position: 5
A; Introns: 53/1; 85/2; 123/3; 177/1.

Query Match 20.4%; Score 70; DB 2; Length 221;
Best Local Similarity 29.6%; Pred. No. 1.4; 16; Indels 16; Gaps 2;
Matches 16; Conservative 7; Mismatches 15; Indels 16; Gaps 2;

Qy 3 ENPACQRQLQSCQEPDDLKQAKCESRTKLEY-----DPRCVYD 41
Db 165 INNCAPKCFSSPDKKDENYLK-CETKCAKLFRTMEEDDAFKSFENCNQACVYD 217

RESULT 6
T78355
hypothetical protein p3 - Mycoplasma hyorhinis
C; Species: Mycoplasma hyorhinis
C; Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 07-Dec-1999
C; Accession: T78355
R; Deng, G.; McIntosh, M.A.
J. Bacteriol. 176, 5929-5937, 1994
A; Title: An amplifiable DNA region from the Mycoplasma hyorhinis genome.
A; Reference number: Z18888; MUID:95014025
A; Accession: T18355
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: DNA
A; Residues: 1-187 <DEN>
A; Cross-references: EMBL:L11447; NID:9150156; PID:9664956; PIDN:AAA62228.1
A; Genetics:
A; Genetic code: SGC3

Query Match 20.1%; Score 69; DB 2; Length 1187;
Best Local Similarity 28.6%; Pred. No. 7.6; 16; Indels 16; Gaps 2;

Matches 20; conservative 12; mismatches 22; indels 16; gaps 5;
QY 1 TENPCQAQRQLQS--CQOPDDIKRQKAC---ESRCTKLEYDPRCV---YDT--GATN 46
Db 403 TEERC--ECTESTGCCEEEACDCSEBECVDETQACLDCTNTQADTKVCGCTQ 460
QY 47 ORHPPGERPR 56
Db 461 EOHTPCEECK 470

RESULT 7
T23681
hypothetical protein M02G9.1 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T23681
R;Matthews, L.
submitted to the EMBL Data Library, November 1996
A; Reference number: Z19781
A; Accession: T23681
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: DNA
A; Residues: 1-513 <WT>
A; Cross-references: EMBL:Z81573; PIDN:CAB04625.1; GSPDB:GN00020; CBSP:M02G9.1
A; Experimental source: clone M02G9
C;Genetics:
A;Gene: CESP::M02G9.1
A;Map position: 2
A;Introns: 23/3; 71/3; 121/3; 183/1; 1083/3; 1141/3; 1408/1; 1449/1
RESULT 8
T31889
hypothetical protein C03A7.8 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T31889
R;Greco, T.; Bradshaw, H.; Elliott, G.
submitted to the EMBL Data Library, July 1997
A; Description: The sequence of *C. elegans* cosmid C03A7.
A; Reference number: Z21096
A; Accession: T31889
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: DNA
A; Residues: 1-438 <GRE>
A; Cross-references: EMBL:AF016451; PIDN:AB65995.1; GSPDB:GN00023; CESP:C03A7.8
A; Experimental source: strain Bristol N2; clone C03A7
C;Genetics:
A;Gene: CESP::C03A7.8
A;Map position: 5
A;Introns: 75/3
C;Superfamily: gliadin

Query Match 20.0%; Score 68.5; DB 2; Length 1513;
Best Local Similarity 28.8%; **Pred.** No. 11; **Mismatches** 11; **Indels** 17; **Gaps** 3;
Matches 15; **Conservative** 9; **Indels** 17; **Gaps** 3;

QY 4 PCQRCQCLSCQOE-----PDDIKQAKCESRCRKLEYDPRCVYDTGAT 45
Db 405 PCQPOCLOCSCLEHQIIPQVVTQLFQCIPO-CQPPAC-----EPQCIOETTT 449

RESULT 8
S08059
alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)
C;Species: *Gossypium hirsutum* (upland cotton)
N;Alternative names: seed storage protein
C;Accession: S08059
R;Chilan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A;Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX.
A;Reference number: S06398
A;Accession: S08059
A;Status: not compared with conceptual translation
A; Molecule type: DNA
A; Residues: 1-509 <CHL>
C;Superfamily: glycinin

Query Match 19.4%; Score 66.5; DB 2; Length 509;
Best Local Similarity 31.6%; **Pred.** No. 6.9; **Mismatches** 10; **Indels** 7; **Gaps** 3;
Matches 18; **Conservative** 10; **Indels** 7; **Gaps** 3;

QY 5 CAQRCQCLSCQOE-----PDDIKQAKCESRCRKLEYDPRCVYDTGATNQ 47
Db 24 CAQRCQCLSCQOE-----PDDIKQAKCESRCRKLEYDPRCVYDTGATNQ 40

RESULT 11
D9423
semaphorin III precursor - human
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-Sep-1999
C;Accession: D4923
R;Kolodkin, A.L.; Matthes, D.J.; Goodman, C.S.
Cell 75, 1389-1399, 1993
A;Title: The Semaphorin genes encode a family of transmembrane and secreted growth co.
A;Reference number: D49423
A; Accession: D94923
A; Status: preliminary; nucleic acid sequence not shown
A; Molecule type: mRNA
A; Residues: 1-771 <K0L>

A;Cross-references: GB:L26081; NID:979932B; PIDN:AAA65938.1; PID:9436560
 C;Genetics:
 A;Gene: GDB:SEMA1
 A;Cross-references: GDB:283448
 C;Superfamily: semaphorin

Query Match 19.4%; Score 66; DB 2; Length 771;
 Best Local Similarity 35.2%; Pred. No. 9.8; Mismatches 23; Indels 7; Gaps 2;
 Matches 19; Conservative 5; Mismatches 23; Indels 7; Gaps 2;

Qy 6 AQRCQLQSQQEPDDLKOKACESRCTKLEYDPRCVYDGTATNORHPPGERTRRQ 59
 Db 510 AQPLPLHRC-----DIYGKACAECC--LARDPVCAMDSACSCSRYPTAKRRRQQ 556

RESULT 12
 T39699 hypothetical protein F31A3.1 - *Caenorhabditis elegans*
 C;Species: *Caenorhabditis elegans*
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
 C;Accession: T29599
 R;Murray, J.; Le, T.
 submitted to the EMBL Data Library, May 1996
 A;Description: The sequence of *C. elegans* cosmid F31A3.
 A;Reference number: 220667
 A;Accession: T29599
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-242 <MUR>
 A;Experimental source: strain Bristol N2; clone F31A3
 C;Genetics:
 A;Gene: CESP:F31A3.1
 A;Map position: X
 A;Introns: 18/3; 160/3
 C;Superfamily: ultra-high-sulfur keratin

Query Match 19.2%; Score 66; DB 2; Length 242;
 Best Local Similarity 43.3%; Pred. No. 4.1; Mismatches 13; Conservative 3; Mismatches 12; Indels 2; Gaps 1;

Qy 2 EMPCAQRCLQSQQEPDDLKOKACESRCTK 31
 Db 175 EPQCCQSCQQCVOQQSMQQ-CASACTK 202

RESULT 13
 T31887 hypothetical protein C03A7.4 - *Caenorhabditis elegans*
 C;Species: *Caenorhabditis elegans*
 C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
 C;Accession: T31887
 R;Graco, T.; Bradshaw, H.; Elliott, G.
 Submitted to the EMBL Data Library, July 1997
 A;Description: The sequence of *C. elegans* cosmid C03A7.
 A;Reference number: Z21096
 A;Accession: T31887
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Residues: 1-388 <GRE>
 A;Molecule type: DNA
 A;Introns: 5/3

Query Match 19.2%; Score 66; DB 2; Length 388;
 Best Local Similarity 43.3%; Pred. No. 6.2; Mismatches 14; Conservative 4; Mismatches 8; Indels 6; Gaps 1;
 Matches 13; Conservative 3; Mismatches 12; Indels 2; Gaps 1;

Qy 5 CAORCLOSC-----QEPDDLKOKACESRCT 30
 Db 317 CAPOCEOSCQOCVQQQPAOCOTACQSCS 348

RESULT 14
 T31888 hypothetical protein C03A7.7 - *Caenorhabditis elegans*
 C;Species: *Caenorhabditis elegans*
 C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
 C;Accession: T31888
 R;Graco, T.; Bradshaw, H.; Elliott, G.
 submitted to the EMBL Data Library, July 1997
 A;Description: The sequence of *C. elegans* cosmid C03A7.
 A;Reference number: Z21096
 A;Accession: T31888
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Residues: 1-388 <GRE>
 A;Experimental source: strain Bristol N2; clone C03A7
 C;Genetics:
 A;Gene: CESP:C03A7.7
 A;Map position: 5
 A;Introns: 7/3
 C;Superfamily: gliadin

Query Match 19.2%; Score 66; DB 2; Length 388;
 Best Local Similarity 43.8%; Pred. No. 6.2; Mismatches 14; Conservative 4; Mismatches 8; Indels 6; Gaps 1;
 Matches 14; Conservative 4; Mismatches 8; Indels 6; Gaps 1;

Qy 5 CAORCLOSC-----QEPDDLKOKACESRCT 30
 Db 317 CAPOCEOSCQOCVQQQPAOCOTACQSCS 348

RESULT 15
 S06398 alpha-globulin type A precursor - upland cotton
 C;Species: *Gossypium hirsutum* (upland cotton)
 C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
 C;Accession: S06398
 R;Chilan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
 Plant Mol. Biol. 9, 533-546, 1987
 A;Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX.
 A;Reference number: S06398
 A;Accession: S06398
 A;Status: not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-605 <CHI>
 C;Superfamily: glycamin
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-605/Product: alpha-globulin type A #status predicted <MAT>
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-388 <GRE>
 A;Cross-references: EMBL:AF016451; PIDN:AAB65001.1; GSPDB:GN00023; CESP:C03A7.4
 A;Experimental source: strain Bristol N2; clone C03A7
 C;Genetics:
 A;Gene: CESP:C03A7.4
 A;Map position: 5
 A;Introns: 75/3
 C;Superfamily: gliadin

Query Match 19.2%; Score 66; DB 2; Length 605;
 Best Local Similarity 38.3%; Pred. No. 9.1; Mismatches 17; Indels 16; Gaps 5;
 Matches 23; Conservative 4; Mismatches 17; Indels 16; Gaps 5;

Qy 9 CLQSC-----QEPDDLK-KACESRCTKLEYDPRCVYDGTATNORHPPGER 54
 Db 102 CEQSCREQVERQQQQPD-QKFECQQRCQWPERKQCVKCREQYQEDWAKER 159

Search completed: March 1, 2001, 15:52:47

Job time: 571 sec

Query Match 19.2%; Score 66; DB 2; Length 388;

Best Local Similarity 43.8%; Pred. No. 6.2; Length 388;

< - >